* Source : Understanding Kalman Filter (1983)

* **Def** : Kalman Filter

KF was invented 1960's, common method for control engineers, physics and statistics people.

KF is a probabilistic model that combines that noisy measurements with the expected trajectory of the object. The main crux of KF is to estimate the value of the "unknown state" based on some observed data (using noisy measurements).

It is called a filtering method since KF finds the best estimate from the noisy data amounts to "filtering out" the noise.

* **Assumptions** :

1) The system must a linear system (data generating process is linear).
2) The noise must follow a gaussian.

- If the system is locally linear systems, we can use Extended KF
- If the system is fully non-linear and non-gaussian, we use Partical KF or Sequential MC.

* **Applications** :

1) Satelite, Drone, GPS Tracking system

2) Weather forecasting : Uses an ensemble KF which could throw out "bad data" that could result in poor forecast.

* **Advantages** : Low cost method (computationally inexpensive)
Handle single / high dimensional data

* **Motivating Example & Formulation of KF** :

Tracking a Satellite location (orbitting around earth). The exact location of the satellite is considered as the "unknown state" which cannot be measured directly. Instead, from tracking stations around the earth, we take measurements of distance to the satellite using sensors (observed / measurment data (noisy)).

We use these measurments to update our estimate of where the satellite truely is.

θ_t : the unknown state / location of satellite at time t (scalar / vector).
Y_t : observed data at time t.

We have this data : $\{Y_t, Y_{t-1}, \dots, Y_1\}$ either scalar or vector.

We assume that Y_t depends on unobservable quantity θ_t. Our aim is to make inference about θ_t whose dimension is indep. of the dim. of Y_t. We assume that the relationship between Y_t and θ_t is linear.

Observation Equation : $Y_t = F_t \theta_t + v_t$, $v_t \sim N(0, \sigma^2_{v_t})$ (1)

where F_t is a known quantity (mapping y_t onto θ_t would be incorporated) and v_t reflects the measurment error with known variance $\sigma^2_{v_t}$.

System Equation : $\theta_t = G_t \theta_{t-1} + w_t$, $w_t \sim N(0, \sigma^2_{w_t})$ (2)

where G_t is a known quantity (G_t would prescribe how the position changes in time according to the physical law governing orbit bodies) and w_t allows for deviation from the physical law owing to non-uniformity of the earth gravitational fields and so on with variance (known) $\sigma^2_{w_t}$.

Note that F_t and G_t may or may not change over time, as it also true for $\sigma^2_{v_t}$ and $\sigma^2_{w_t}$. Finally, not that v_t and w_t are independent.

KF refers to a recursive procedure for inference about the state of nature θ_t. Given the data $\underset{\sim}{Y_t} = \{Y_t, Y_{t-1}, ..., Y_1\}$, we carry out inference about θ_t using Bayesian approach.

Recursive estimation procedure :

$\mathbb{P}$(state of nature | observed data) $\propto$ $\mathbb{P}$(obs data | state) x Prior dist of θ_t

$\underset{\text{posterior}}{\downarrow}$ $\underset{\text{likelihood}}{\downarrow}$

Bayes Theorem: $\mathbb{P}(\theta_t \mid \underset{\sim}{Y_t}) \propto \mathbb{P}(\underset{\text{current data}}{\underset{\downarrow}{Y_t}} \mid \theta_t, \underset{\sim}{Y_{t-1}}) \cdot \mathbb{P}(\theta_t \mid \underset{\sim}{Y_{t-1}})$ (3)

we are conditioning on $\underset{\sim}{Y_{t-1}}$ i.e we observed data upto Y_{t-1}.

We start with finding the distribution of $\theta_t \mid Y_{t-1}, ..., Y_1$. At time $t-1$, our state of knowledge about θ_{t-1} is embodied in the following prob. statement for θ_{t-1} $(\theta_{t-1} \mid \underset{\sim}{Y_{t-1}}) \sim N(\hat{\theta}_{t-1}, \Sigma_{t-1})$ (4) where $\hat{\theta}_{t-1}$ and Σ_{t-1} are expectation and variance respectively.

At time 0, we can choose $\hat{\theta}_0$ and Σ_0 to be our best guesses about mean and variance of θ_0 (some constant). We will work on time t

in two stages.

Stage 1: Prior to observing Y_t

Conditioning (2) on $\underset{\sim}{Y}_{t-1}$, we get:

$$\theta_t \mid \underset{\sim}{Y}_{t-1} = \overset{\overset{G_t \text{ is known}}{\wedge}}{(G_t \mid \underset{\sim}{Y}_{t-1})} (\theta_{t-1} \mid \underset{\sim}{Y}_{t-1}) + (w_t \mid \underset{\sim}{Y}_{t-1}) \rightarrow \text{indep of the data}$$

$$= G_t \cdot (\theta_{t-1} \mid \underset{\sim}{Y}_{t-1}) + w_t$$

$$\mathbb{E}(\theta_t \mid Y_{t-1}) = \mathbb{E}(G_t \cdot (\theta_{t-1} \mid \underset{\sim}{Y}_{t-1}) + w_t)$$

$$= G_t \, \mathbb{E}(\theta_{t-1} \mid \underset{\sim}{Y}_{t-1})$$

$$= G_t \cdot \hat{\theta}_{t-1}$$

$$Var(\theta_t \mid Y_{t-1}) = Var(G_t (\theta_{t-1} \mid \underset{\sim}{Y}_{t-1}) + w_t)$$

$$= G_t \cdot G_t^T \, Var(\theta_{t-1} \mid Y_{t-1}) + \sigma^2_{w_t}$$

$$= G_t \, G_t^T \cdot \Sigma_{t-1} + \sigma^2_{w_t} = R_t$$

Hence, $\theta_t \mid \underset{\sim}{Y}_{t-1} \sim N(G_t \hat{\theta}_{t-1}, \; G_{t-1} \Sigma_{t-1} G^T + \sigma^2_{w_t} = R_t)$ (5)

* Stage 2: After observing Y_t

We have to compute $\mathbb{P}(Y_t \mid \theta_t, \underset{\sim}{Y}_{t-1})$

We start with defining the error e_t in predicting Y_t as follows:

$$e_t = Y_t - \hat{Y}_t$$

$$= Y_t - \mathbb{E}(Y_t \mid \underset{\sim}{Y}_{t-1})$$

$$= Y_t - F_t \, \mathbb{E}(\theta_t \mid Y_{t-1}) \qquad \text{by } 1$$

$$= Y_t - F_t \cdot G_t \, \hat{\theta}_{t-1} \qquad \text{by } 5$$

Error: $e_t = Y_t - \underbrace{F_t \cdot G_t}_{\text{know quantities}} \cdot \hat{\theta}_{t-1}$ (6)

Observing Y_t is equivalent to observing e_t ($Y_t = e_t + F_t \cdot G_t \cdot \hat{\theta}_{t-1}$)

We can rewrite 3 : $\mathbb{P}(\theta_t | e_t, \underset{\sim}{Y_{t-1}}) \propto \mathbb{P}(e_t | \theta_t, \underset{\sim}{Y_{t-1}}) \, \mathbb{P}(\theta | \underset{\sim}{Y_{t-1}})$ (7)

this is the updated posterior $\quad$ we know this

Now, we want to find : $\mathbb{P}(e_t | \theta_t, \underset{\sim}{Y_{t-1}})$

$$
\begin{aligned}
\mathbb{E}(e_t | \theta_t, \underset{\sim}{Y_{t-1}}) &= \mathbb{E}(e_t) \qquad \text{since } \theta_t \text{ is observed} \\
&= \mathbb{E}(Y_t - F_t G_t \hat{\theta}_{t-1}) \\
&= \mathbb{E}(F_t \theta_t + v_t - F_t G_t \hat{\theta}_{t-1}) \\
&= F_t \theta_t - F_t G_t \hat{\theta}_{t-1}
\end{aligned}
$$

$$
\begin{aligned}
Var(e_t | \theta_t, \underset{\sim}{Y_{t-1}}) &= Var(e_t) \\
&= Var(Y_t - F_t G_t \hat{\theta}_{t-1}) \\
&= Var(F_t \theta_t + v_t - F_t G_t \hat{\theta}_{t-1}) \\
&= Var(v_t) \\
&= \sigma^2_{v_t}
\end{aligned}
$$

$$e_t | \theta_t, Y_{t-1} \sim N(F_t \theta_t - G_t F_t \hat{\theta}_{t-1}, \sigma^2_{v_t}) \qquad (8)$$

Recap : Bivariate Normal Distribution

X_1 and X_2 jointly follow BN distribution with mean μ_1, μ_2 with covariances $\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

1) $X_1 \sim N(\mu_1, \Sigma_{11}) \rightarrow 8$

$X_2 \sim N(\mu_2, \Sigma_{22}) \rightarrow 5$

$$X_1 \mid X_2 \sim N(\mu_1 + \underbrace{\Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)}, \; \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \qquad (9)$$

works like a regression function
with $\Sigma_{12}\Sigma_{22}^{-1}$ as the coefficient
of regression $\frac{S_{XY}}{S_{XX}}$

$X_1 \hat{=} e_t \qquad X_2 \hat{=} \theta_t$

corresponds to

By 5, $\mu_2 \hat{=} G_t . \hat{\theta}_{t-1}$ and $\Sigma_{22} \hat{=} R_t$

$$e_t \mid \theta_t, \underset{\sim}{Y_{t-1}} \sim N(\mu_1 + \Sigma_{12} . R_t^{-1}(\theta_2 - G_t . \hat{\theta}_{t-1}), \Sigma_{11} - \Sigma_{12} R_t^{-1} . \Sigma_{21}) \quad (10)$$

We will equate 8 and 10 :

$$\mu_1 + \Sigma_{12} R_t^{-1}(\theta_t - G_t \hat{\theta}_{t-1}) = F_t(\theta_t - G_t \hat{\theta}_{t-1}) \Rightarrow \mu_1 = 0,$$

$$\Sigma_{12} R_t^{-1} = F_t$$

$$\Rightarrow \Sigma_{12} = F_t . R_t$$

$$\Sigma_{11} - \Sigma_{12} R_t^{-1} . \Sigma_{21} = \sigma_{v_t}^2 \Leftrightarrow \Sigma_{11} - F_t R_t . R_t^{-1} . R_t . F_t = \sigma_{v_t}^2$$

$$\Leftrightarrow \Sigma_{11} = F_t . R_t F_t^T + \sigma_{vt}^2$$

By the converse relationship of BN :

$$\begin{pmatrix} e_t \\ \theta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ G_t \hat{\theta}_{t-1} \end{pmatrix}, \begin{pmatrix} F_t . R_t . F_t^T + \sigma_{vt}^2 & F_t . R_t \\ R_t^T F_t & R_t \end{pmatrix} \right)$$

$$\left[\begin{pmatrix} \theta_t \\ \end{pmatrix} \middle| Y_{t-1} \right] \sim N\left(\begin{pmatrix} G_t \hat{\theta}_{t-1} \\ \end{pmatrix}, \begin{pmatrix} R_t & R_t F_t^T \\ \end{pmatrix} \right) \quad (11)$$

$$\left[\begin{pmatrix} \theta_t \\ e_t \end{pmatrix} \middle| Y_{t-1} \right] \sim N\left[\begin{pmatrix} G_t \theta_{t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} R_t & R_t F_t^T \\ F_t R_t & F_t R_t F_t^T + \sigma^2_{vt} \end{pmatrix} \right] \quad (11)$$

$$\theta_t \mid e_t, Y_{t-1} \sim N\left(\overbrace{G_t \hat{\theta}_{t-1} + F_t R_t . (F_t R_t F_t + \sigma^2_{vt})^{-1} e_t}^{\hat{\theta}_t} \; , \; \underbrace{R_t - R_t . F_t^T}_{\Sigma_t} \right)$$

$$(F_t R_t F_t^T + \sigma^2_{vt}) F_t R_t$$

Posterior Mean : $\hat{\theta}_t$

Posterior Variance: Σ_t

* A simple KF Model : Obs Eq. : $Y_t = \theta_t + v_t \quad , \quad v_t \sim N(0, \sigma^2_{vt})$

Transition : $\theta_t = \theta_{t-1} + w_t \quad , \quad \theta_t \sim N(0, \sigma^2_{wt})$

This model corresponds to a class of ARIMA(0,1,1) model.

$$Y_t = \theta_t + v_t \quad ; \quad v_t \sim N(0, 2)$$

$$\theta_t = \theta_{t-1} + w_t \quad ; \quad w_t \sim N(0, 1)$$

This corresponds to a simple exponential smoothing model.